$c = c_0$, then one would get a distinct equilibrium position in the range $0 < \theta < \pi$.

This example is of interest in the instruction of elementary mechanics since it is a more complicated system that exhibits neutral stability than the usual example of a smooth uniform sphere lying on a flat horizontal surface.

Einstein's first derivation of mass–energy equivalence

John Stachel[a)]
Einstein Project, Institute of Advanced Study
and Physics Department, Princeton University, Princeton, New Jersey 08540
Roberto Torretti
Departamento de Filosofía, Universidad de Puerto Rico, Río Piedras, Puerto Rico 00931

(Received 24 April 1981; accepted for publication 13 August 1981)

There is a persistent tradition, apparently starting with a paper by Ives, that Einstein's original derivation of the mass–energy equivalence was fallacious.[1] It has probably gained most currency through the account in Jammer's well-known book *Concepts of Mass.*[2] Jammer states[3]:

> It is a curious incident in the history of scientific thought that Einstein's own derivation of the formula $E = mc^2$, as published in his article in the *Annalen der Physik*, was basically fallacious. In fact, what for the layman is known as "the most famous mathematical formula ever projected"[4] in science was but the result of a *petitio principii*, the conclusion of begging the question.

Arzeliès, in his treatise on relativity,[5] also adopts Ives's discussion. He states[6]:

> It might appear rather piquant that a relation of this importance was introduced into physics by this expedient. In fact it is just one example among others of the slight importance of logic in physical research.

We leave a decision about the importance of logic in writing treatises on relativity to the reader after finishing this article.

Arthur Miller, in his recent book on the emergence and early interpretation of special relativity,[7] endorses Jammer's verdict.

Before considering the quality of the arguments on which Ives, Jammer, and Arzeliès base their criticism, let us repeat Einstein's derivation. As we shall see, far from being "basically fallacious," it is quite sound.

Einstein[8] considers a body at rest in an inertial frame which emits electromagnetic radiation (Einstein speaks of a light wave) of total energy L in two equal but oppositely directed amounts. Since the emissions are symmetrical, the body loses energy but not momentum, so it remains at rest in the original inertial system. Einstein assumes that its initial energy with respect to this frame was E_0; and that energy is conserved in this frame, so that its final energy E_1 with respect to the frame differs from E_0 by the amount L:

$$E_0 = E_1 + L. \tag{1}$$

Now we consider the same act of emission from another inertial frame, with respect to which the body is moving with a velocity **v**. Note that **v** is arbitrary, and that its magnitude $|v|$ can be chosen as small as we desire. If energy conservation is assumed in the first inertial frame, the relativity principle requires it to hold in the second as well. If H_0 and H_1 are the initial and final energies with respect to the second frame, Einstein shows that the initial and final energies differ by the amount $L/\surd(1 - v^2/c^2)$:

$$H_0 = H_1 + L \surd(1 - v^2/c^2). \tag{2}$$

This result is established by using the law for the transformation of the energy of a plane light wave from one inertial frame to the other, derived in his first paper on the special theory of relativity.[9] Now Einstein subtracts Eq. (1) from Eq. (2) to obtain:

$$(H_0 - E_0) - (H_1 - E_1) = L\,[1/\surd(1 - v^2/c^2) - 1]. \tag{3}$$

Einstein proceeds:

> H and E are energy values of the same body, referred to two coordinate systems in motion relative to each other, in one of which [...] the body is at rest. *It is therefore clear* that the difference $H - E$ can differ from the kinetic energy K of the body with respect to the other system [...] only by an additive constant C which depends on the choice of the arbitrary additive constants in the energies H and E.[10]

We may therefore set

$$\begin{aligned} H_0 - E_0 &= K_0 + C, \\ H_1 - E_1 &= K_1 + C, \end{aligned} \tag{4}$$

for the constant C will not be altered by the emission of radiation. Substituting from Eq. (4) into Eq. (3), we obtain

$$K_0 - K_1 = L\,[1/\surd(1 - v^2/c^2) - 1], \tag{5}$$

where K_0 and K_1 are the initial and final kinetic energies of the body with respect to the inertial frame in which its velocity is **v**. This, the decisive step of Einstein's derivation, depends on the acceptance of Eqs. (4). Though "clear" to Einstein, their validity has not been quite so evident to others. The following relativistic argument will, we hope, contribute to make it so.

Consider an isolated body in equilibrium, and therefore at rest in some inertial frame, which we call its rest frame. Its internal state will be characterized by a set of parameters whose values (with respect to its rest frame) we symbolize collectively by S. By the relativity principle, it must be possible for the body to have the same internal state S when in uniform motion with velocity **v**. For otherwise one could determine the absolute state of motion of the body by determining its internal state. Its energy in its rest frame (in the absence of any external fields, of course) is a function of

© 1982 American Association of Physics Teachers

its internal state only. Then its energy in an inertial frame with respect to which it is moving can only be a function of S and its velocity. Because of the isotropy and homogeneity of space, the body's energy can depend only on the magnitude of its velocity—as well as on its internal state:

$$E = E(v,S).$$

The kinetic energy of the body, by definition, is equal to the work necessary to bring the body from the state of rest to uniform motion with velocity $\mathbf{v}$. But, by conservation of energy, this must be equal to the difference between its energy for the state S and speed v and its energy for the same internal state when at rest:

$$K = E(v,S) - E(0,S).$$

K depends on v and S:

$$K = K(v,S).$$

We can now rewrite Eqs. (1)–(5) in our new notation, by setting

$$E_0 = E(0,S_0), \quad H_0 = E(v,S_0), \quad K_0 = K(v,S_0);$$
$$E_1 = E(0,S_1), \quad H_1 = E(v,S_1), \quad K_1 = K(v,S_1),$$

where S_0 and S_1 denote, respectively, the initial and the final internal state of the body, assumed to settle into a new equilibrium state S_1 after emitting the radiation. Obviously,

$$[E(v,S_0) - E(0,S_0)] - [E(v,S_1) - E(0,S_1)] = K(v,S_0) - K(v,S_1),$$

whence Eq. (5), that is, in our notation,

$$K(v,S_0) - K(v,S_1) = L\,[1/\surd(1 - v^2/c^2) - 1], \tag{5'}$$

follows at once.

Note how our argument rests on the relativity principle. We have proceeded from a consideration of the same body with respect to two inertial frames to a comparison of the body in two states of motion with respect to the same inertial frame; in other words, from a passive to an active interpretation of the Lorentz transformation. But the equivalence of the two viewpoints is just what the relativity principle (indeed any symmetry principle) asserts: the active interpretation is always possible.[11]

Einstein remarks that Eq. (5) means that "the kinetic energy of the body [...] decreases as a consequence of the emission of light, and indeed by an amount independent of the qualities of the body." This is not strictly accurate, since the amount does depend somehow on the relationship between S_0 and S_1; but what Einstein presumably had in mind is expressed in our notation by the fact that the right-hand side of the equation is independent of S_0 and S_1. He immediately notes that "The difference [of the two kinetic energies] depends on the velocity in the same way as the kinetic energy of the electron," which he had previously calculated in his first paper on special relativity.[12] Note that this is a *result* of his argument, not an assumption or a trivial consequence of his having assumed that $K(v,S)$ for any body was of the same form as the kinetic energy of the electron.

Finally, Einstein draws his conclusion from Eq. (5) by passing to the Newtonian limit:

> Neglecting magnitudes of fourth and higher order [in v/c] we obtain
>
> $$K(v,S_0) - K(v,S_1) = (1/2)(L/c^2)v^2.$$
>
> From *this* equation it immediately follows: If a body emits energy L in the form of radiation, then its mass diminishes by L/c^2.[13]

This terse and subtle text demands some comment. Passage to the Newtonian limit is justified and does not impair the exact validity of the conclusion,[14] because, as we have already noted, the moving frame, which is introduced only for the sake of comparison, can be arbitrarily slow. Einstein is arguing that a body *at rest*, which emits energy in such a way as to remain at rest, thereby decreases its inertial mass. He is not directly concerned with how the energy of the body varies with its velocity. The reader should also note how carefully Einstein avoids any assumption about how the relativistic kinetic energy depends on velocity, mass, or any other parameters of the body's internal state.

We can define the inertial mass[15] for a body in translational motion (in keeping with the requirement that special relativistic dynamics have a Newtonian limit as $v \to 0$) by

$$m = \lim_{v \to 0} \frac{K}{v^2/2}.$$

It follows from Eq. (6) that

$$\lim_{v \to 0} \frac{(K_0 - K_1)}{v^2/2} = \frac{L}{c^2}.$$

With the above definition of inertial mass, this can be rewritten

$$m_0 - m_1 = L/c^2,$$

which is clearly an exact, not an approximate, result.

We are therefore entitled to the inference that a body at rest emitting electromagnetic radiation must lose inertial mass equal in amount to the energy lost divided by c^2.

Einstein then remarks that "it is obviously inessential that the energy subtracted from the body should turn precisely into energy of radiation." He then jumps to the conclusion that "the mass of a body is a measure of its energy content," which clearly suggests that the rest mass of a body might be totally convertible into other forms of energy. Soon after, Einstein sought to substantiate the above remark by proving the mass–energy equation for other cases of energy exchange due to electromagnetic interaction, but not involving the emission or absorption of radiation.[16] However, it was not until much later that he gave a purely dynamic derivation of the mass–energy equivalence.[17] The final conclusion: that the entire mass of a body is in effect a measure of its energy, is of course unwarranted by Einstein's premises. This is what he wrote in defense of it in 1907:

> A mass μ is equivalent—insofar as its inertia is concerned—to an energy content of the magnitude μc^2. Since we can arbitrarily fix the zero of [the rest energy], we are not even able to distinguish, without arbitrariness, between a "true" and an "apparent" mass of the system. It appears much more natural to regard all inertial mass as a store of energy.[18]

Later events have obviously borne out this presumption.

We turn now to the consideration of Ives's, Jammer's, and Arzeliès's criticism of Einstein's argument.[19] None of the critics disputes that up to Eq. (3) it is correct, but they all take exception to Einstein's next step. To Ives "it is by no means 'clear that, etc.' " in the Einstein passage quoted above right after Eq. (3). Jammer voices an even stronger disapproval. According to him, after "correctly" proving Eq. (3), Einstein "mistakenly put" $H - E$ [or, in our notation, $E(v,S) - E(0,S)$] equal to the kinetic energy. And yet it is hard to see what else one could mean by the kinetic energy of a body with internal state S and speed v. Ives, Jammer, and Arzeliès do not tell us why they believe that Eqs. (4) are doubtful or erroneous, but immediately embark on the following remarkable exercise in logical analysis, for which Jammer and Arzeliès give all due credit to Ives. They assume that the kinetic energy of a body of rest mass m, with respect to an inertial frame in which it is moving with speed v is equal to

$$K = mc^2[1/\sqrt{(1 - v^2/c^2)} - 1]. \qquad (6)$$

Einstein had proved Eq. (6) for the kinetic energy of an "electron," i.e., a charged structureless particle, in his first relativity paper[20]; but he studiously avoided using it in the derivation of the mass–energy equivalence, even though, as we shall now see, it would have simplified his task. (Had he used it he could indeed have been justly accused of question begging, for he had as yet no grounds for assuming that the dependence of the kinetic energy on internal parameters can be summed up in a rest mass term. Only if we know the mass–energy equivalence relation will this seem reasonable.) From Eqs. (6) and (3), Ives, Jammer and Arzeliès infer that

$$(H_0 - E_0) - (H_1 - E_1) = [L/(m_0 - m_1)c^2](K_0 - K_1), \qquad (7)$$

where we denote the vacuum speed of light by c, the initial and final rest mass of the body by m_0 and m_1, respectively, and otherwise use Einstein's notation. Comparing Eq. (7) and Eqs. (3) and (5), we verify that, if Eq. (6) is assumed, Eq. (5) holds good if and only if

$$m_0 - m_1 = L/c^2. \qquad (8)$$

Thereupon one would normally conclude that, if Eq. (6) is added to Einstein's premises, the mass–energy relation (8) follows directly from them, without having to go to the Newtonian limit. For surely, if Eq. (5) is true only if Eq. (8) is true, and Einstein's premises entail that Eq. (5) is true, they also entail the truth of Eq. (8), as a single application of *modus ponendo ponens* will make clear. At this point, however, Ives, Jammer, and Arzeliès veer off the trodden path of ordinary propositional logic and pronounce their indictment. According to Ives, "what Einstein did by setting down [Eq. (4)] as 'clear' was to *introduce* the relation [Eq. (8)]. *Now this is the very relation the derivation was supposed to yield.*" According to Jammer, by comparing Eq. (7) with Einstein's assumption (4), "we see that Einstein unwittingly assumed" Eq. (8), "which is exactly the contention to be proved." All three authors agree that Einstein has thereby committed a "fallacy," which Jammer, as we saw, diagnoses as a *petitio principii*. Arzeliès similarly speaks of a "vicious circle." Now it is plain that Eqs. (6) and (8) jointly imply that Eq. (3) entails Eq. (5). If Einstein had used this method for proving Eq. (5) and had then proceeded to infer Eq. (8) from it, he would indeed have perpetrated a *petitio*. The method, however, does not occur in his work, but is a "rational reconstruction" which we have just devised in order to make sense out of the claims of his critics. (It is the method that he ought to have followed to earn their reprimand.) Nowhere in his paper does he invoke Eq. (6), let alone Eq. (8), to prove Eq. (5), which, as we saw, he derives from Eqs. (3) and (4). Jammer asserts without proof that Eqs. (4) are "mistaken," and if he were right Einstein's derivation would indeed break down. Not, be it noted, because if Eqs. (4) are wrong Einstein's argument would, by the accepted standards, be fallacious; but rather because, though logically impeccable, it would then rest on a falsehood. In all likelihood, Ives, Jammer, and Arzeliès are innocent of the gross misrepresentation of Einstein's method of proof which we have put forward on their behalf. If that is the case, all they can possibly be objecting to is the fact that the goal of Einstein's argument, i.e., Eq. (8), is a necessary condition of the joint assertibility of Einstein's premises—so that the latter cannot be true unless the former is true as well. But this is a relation in which the conclusion of an argument should stand to its premises whenever the argument is valid by standard—and by every known nonstandard—logic. Thus if we are not willing to countenance some mind-boggling metalogical innovation, we have to declare that Ives, Jammer, and Arzeliès—not Einstein—are guilty of a logical error.

Ives seems to have been inspired in his criticism by a remark of Planck's, which he quotes. Commenting on Einstein's derivation in a 1907 paper, Planck[21] said that it was based on an "assumption permissible only as a first approximation." To understand Planck's point, made in a footnote near the end of the paper, we must go back to its Introduction. He states there:

> We usually conceive the total *energy* of a moving ponderable body as additively composed of a term which is independent of the internal state of the body and only varies with its velocity: the translational energy; and a second term which is independent of the velocity and only depends on the internal state [...]: the internal energy of the body.[22]

He then points out that any ponderable body contains a contribution to its energy from thermal radiation, and that the thermal radiation energy does not fall into two such terms. Therefore "a division of the energy into internal and translational energy" is not possible.

Thus after he has established the mass–energy equivalence for thermal energy, he remarks in a footnote:

> Essentially the same conclusion was drawn by A. Einstein [...] from the application of the relativity principle to a particular radiation process, however, with the assumption, permissible only as a first approximation, that the total energy of a moving body is additively composed of its kinetic energy and its energy with respect to a system in which it is at rest.[23]

But, as we have seen above, the definition of the kinetic energy implicit in Einstein's derivation, and which we have made explicit, does not at all imply that the kinetic energy is "independent of the internal state of the body." All that is implied by our definition is that the net energy necessary to bring the body from an initial state of rest in internal state S to a final state of motion with velocity $\mathbf{v}$ and the same internal state S must be independent of the particular process by which it was brought to that final state. But this is just the principle of energy conservation.

To summarize: Einstein uses the following principles in his derivation: the principle of relativity, the law of conservation of energy, the existence of a Newtonian limit for relativistic dynamics, and the relativistic law of transformation of the energy of an electromagnetic wave. These premises are certainly strong enough to derive the mass–energy equivalence relation; and in that sense, of course, the conclusion is required by the premises, as indeed it must be in any logically correct argument. If one of his assumptions turned out to be wrong, that would make his argument inconclusive, but not fallacious.

---

[a] On leave from Dept. of Phys., Boston University.

[1] H. I. Ives, J. Opt. Soc. Am. **42**, 540 (1952). Ives's criticism of Einstein's derivation was criticized in turn by J. Riseman and I. G. Young, J. Opt. Soc. Am. **43**, 618 (1953), to whom Ives replied in the same issue, p. 619. This exchange did not touch on the issues discussed in the present ariticle.

[2] M. Jammer, *Concepts of Mass in Classical and Modern Physics* (Harvard University, Cambridge, MA, 1961).

[3] Reference 2, p. 177.

[4] Jammer here quotes William Cahn, *Einstein* (Citadel, New York, 1955).

[5] H. Arzeliès, "Études Relativistes;" the discussion of Einstein's various proofs of $E = mc^2$ is in the volume *Rayonnement et dynamique du corpuscule chargé fortement acceléré* (Gauthier-Villars, Paris, 1966), pp. 74–79.

[6] Reference 5, p. 76.

[7] A. I. Miller, *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905–1911)* (Addison-Wesley, Reading, MA, 1981). On p. 377 Miller states: "Ives' [Ref. 1] analysis of Einstein's [Ref. 8] derivation of the mass–energy equivalence revealed a logical inconsistency—Einstein had assumed the result to be proven at the outset."

[8] A. Einstein, Ann. Phys. Ser. 4 **18**, 639 (1905). An English translation may be consulted in *The Principle of Relativity* (Methuen, London, 1923—subsequently reprinted by Dover), pp. 67–71; but, since the translation is sometimes not quite accurate, we have translated directly from the German. We use the standard symbol c, instead of Einstein's v, to designate the speed of light *in vacuo*.

[9] A. Einstein, Ann. Phys. Ser. 4 **17**, 891 (1905). The proof is given in Sec. 8. See Ref. 8 for reference to one English translation in *The Principle of Relativity*. A better translation is contained in the Appendix to Miller, Ref. 7.

[10] Einstein, Ref. 8, pp. 640ff. Emphasis added.

[11] For those to whom the talk of active and passive transformations sounds esoteric, the following consideration may be helpful. Let the body with internal state S be initially at rest in an inertial frame F and moving with velocity $\mathbf{v}$ in an inertia frame F'. Its energy with respect to F is then $E(0,S)$. Now accelerate the body to velocity $-\mathbf{v}$ in F. According to what we said above, its energy with respect to F increases thereby to $E(v,S) = E(0,S) + K(v,S)$—where $v = |\mathbf{v}|$. Since the body with internal state S is now at rest in F', its energy with respect to F' will be, by the relativity principle, equal to $E(0,S)$. Returning to the initial situation we see that, by symmetry, when the body had energy $E(0,S)$ with respect to F, it must have had—*modulo* an arbitrary constant independent of v and S—the energy $E(v,S) = E(0,S) + K(v,S)$ with respect to F'.

[12] Einstein, Ref. 9, p. 920 (Sec. 10).

[13] Einstein, Ref. 8, p. 641. Emphasis added. We have also translated Einstein's notation into ours.

[14] We cannot agree with Whittaker's statement that the proof "was put forward only as approximate." See E. T. Whittaker, *A History of the Theories of Aether and Electricity, Vol. II, The Modern Theories 1900–1926* (Nelson, London, 1953), p. 52. Whittaker also notes without comment that Einstein's "reasoning has been criticised" by Ives.

[15] Although a variable inertial mass, and even transverse and longitudinal masses were used in the early days of SRT, Einstein came to the view that only one (rest) mass should be defined; and that otherwise one should rather speak about the law of variation of energy or momentum with velocity.

[16] A. Einstein, Ann. Phys. Ser. 4 **23**, 371 (1907); Jahrb. Radioaktivität **4**, 411 (1907); **4**, 440ff (1907).

[17] A. Einstein, Am. Math. Soc. Bull. **41**, 223 (1935). As is well known, relativistic dynamics, free from electromagnetic presuppositions, was founded by G. N. Lewis and R. C. Tolman, Philos. Mag. **18**, 510 (1909).

[18] Einstein, Ref. 16, second article, p. 442.

[19] Ives, Ref. 1, pp. 542–543; Jammer, Ref. 2, pp. 177–179.

[20] Einstein, Ref. 9, pp. 919 ff.

[21] M. Planck, K. Preuss. Akad. Wiss. Sitzungsber. (2), 542 (1907).

[22] Planck, Ref. 21, p. 542.

[23] Planck, Ref. 21, p. 566.

Erratum: "Quantum solutions for a symmetric double square well" [Am. J. Phys. 50, 239 (1982)]

Edward A. Johnson and H. Thomas Williams

Figures 2–16 were reproduced upside down. The editors apologize for this mistake.